115 ways not to say “Hello, World!”

Syntax errors observed in a large-scale online CS0 Python course

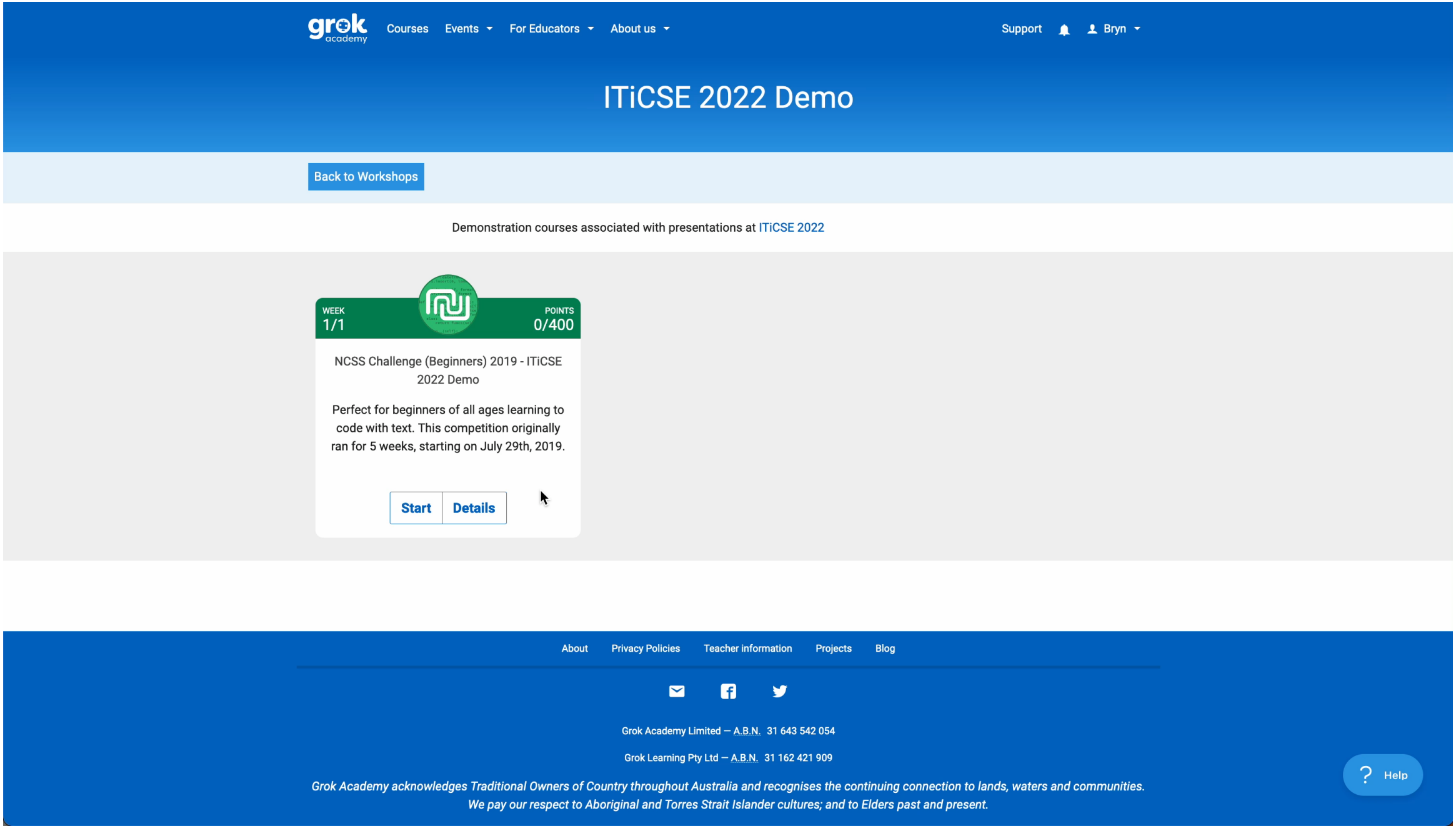
Bryn Jeffries

Irena Koprinska
Jung A. (Monica) Lee

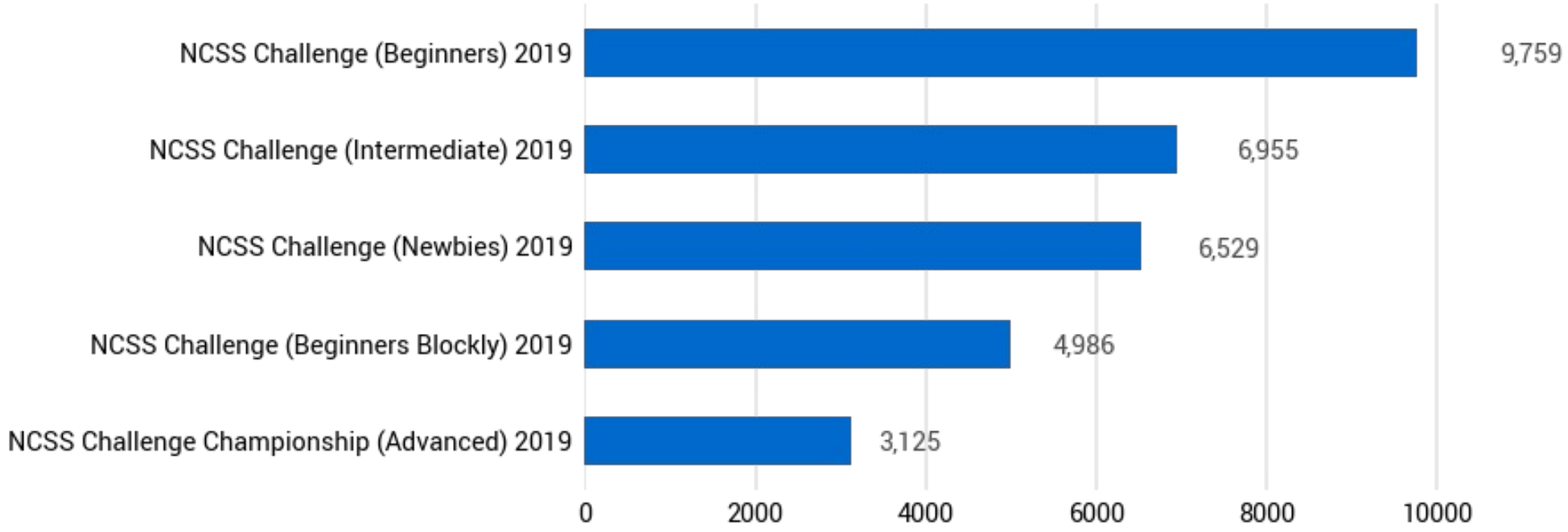
Australian National Computer Science School (NCSS) Challenge

- https://grokacademy.org/challenge/
- 5-week programming course (next one starts on July 25)
- Aligned to Australian federal Digital Technologies curriculum
- Multiple streams at different difficulties
  - Newbies (Blockly)
  - Beginners (Blockly and Python options)
  - Intermediate (Python only)
  - Advanced (Python only)
- Supported by ~100 online tutors
  - see Jeffries et al., ACE'20 10.1145/3373165.3373172

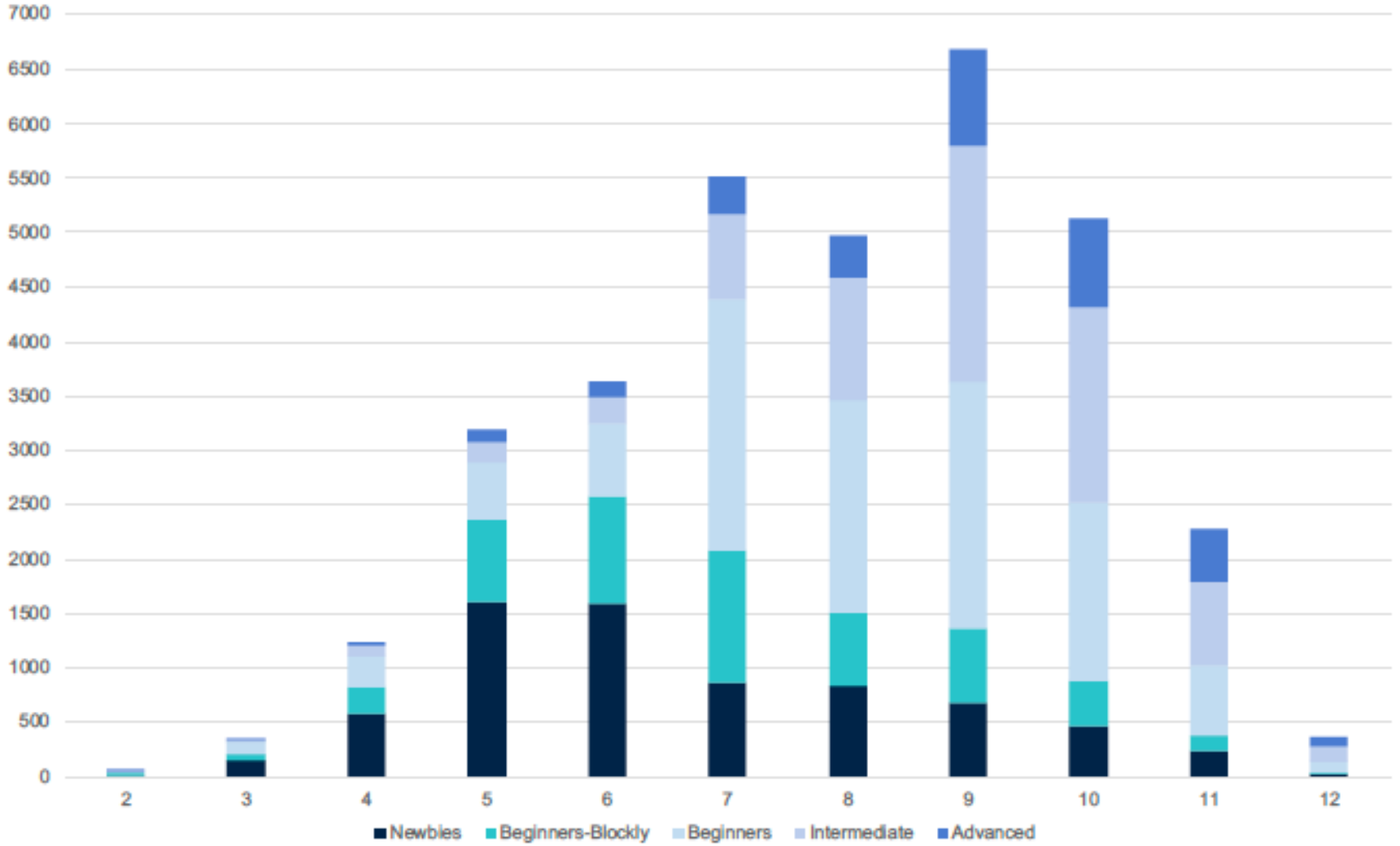
https://grok.ac/iticse2022



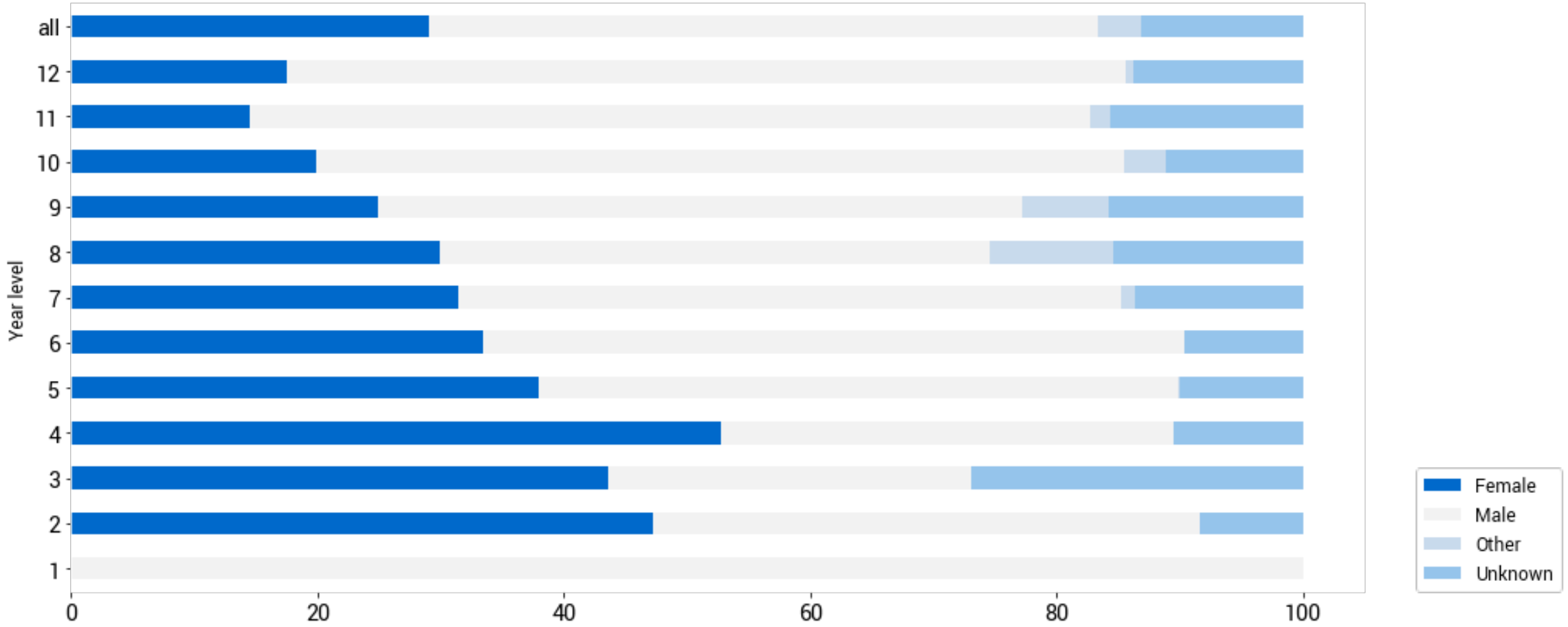
Australian enrolled students by stream



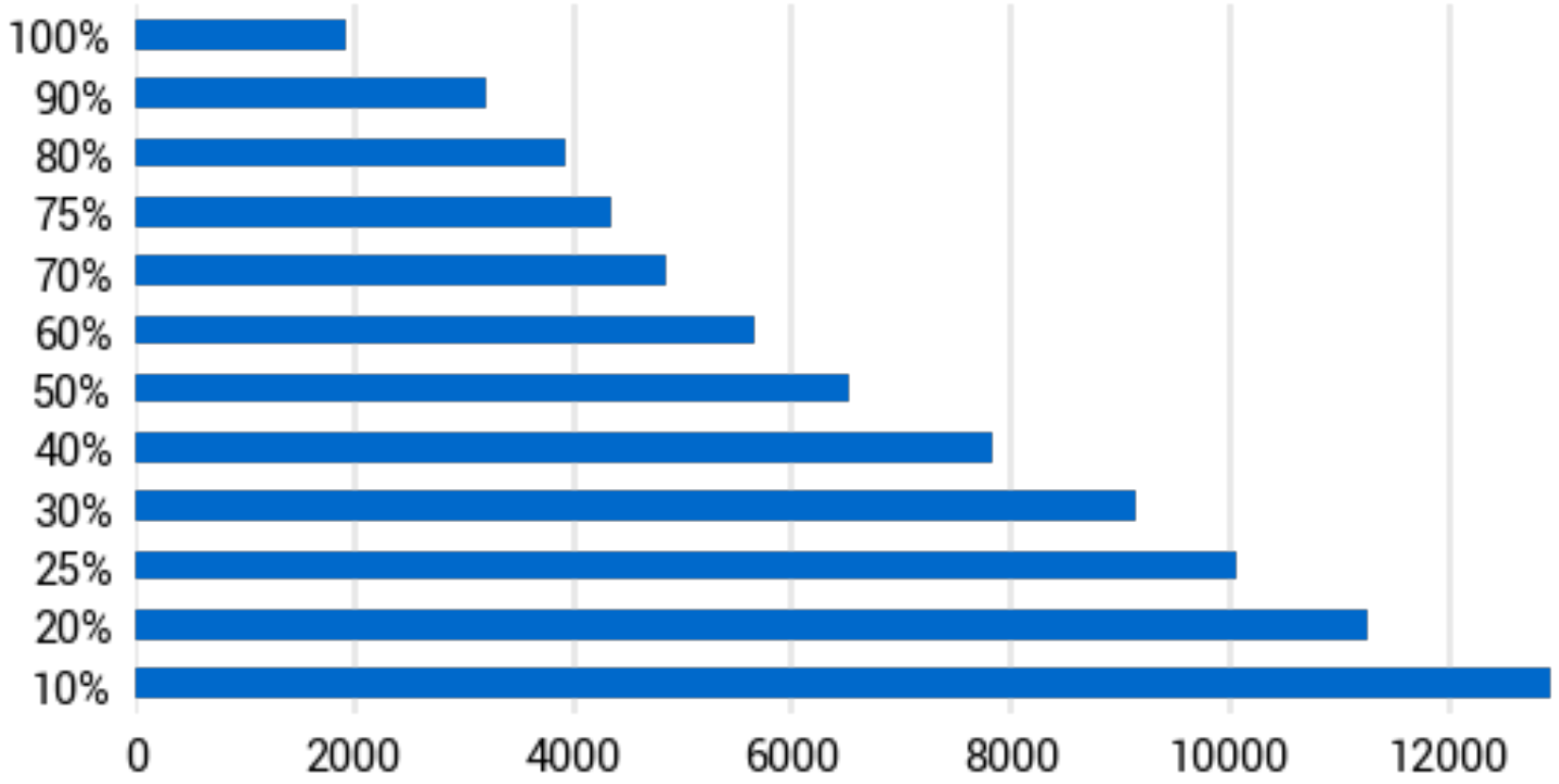
Australian student enrolment by school year



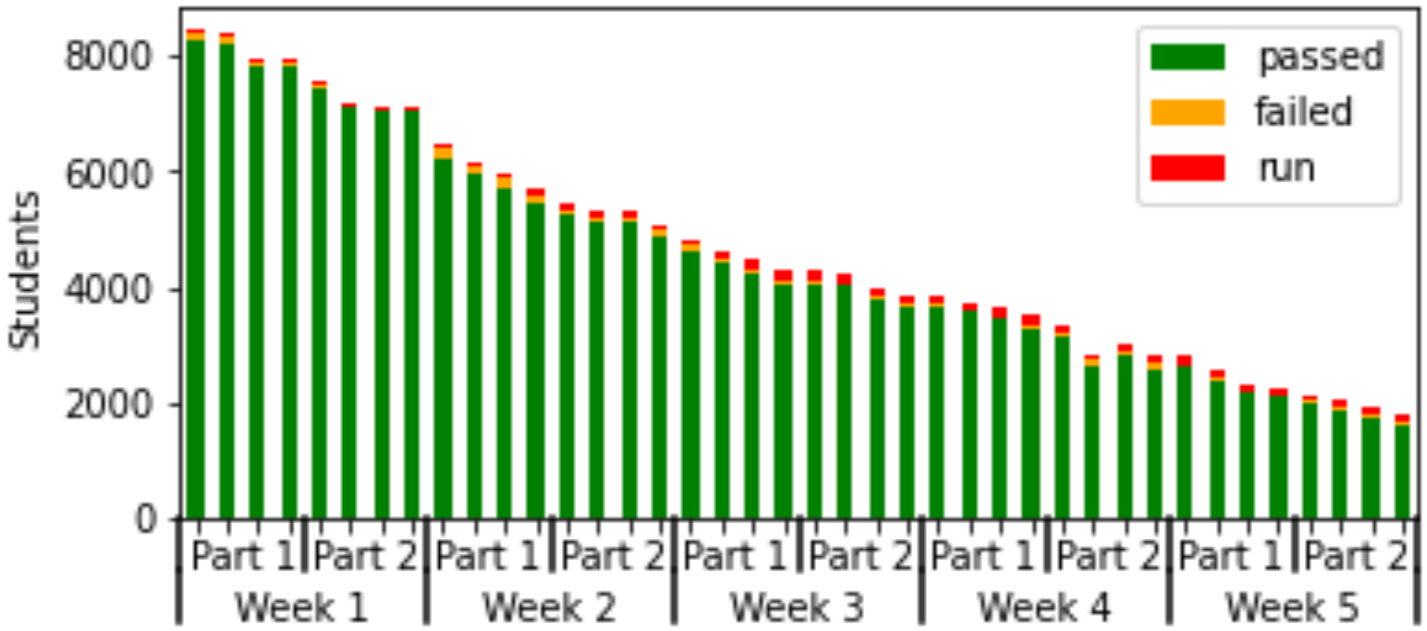
Gender Distribution



Student course completion rates



Beginners (Python) Challenge problem outcomes



Markov Chain analysis [Jeffries et al., ACE'20 10.1145/3373165.3373172]:

- 16% of students failing a problem will skip the next
- 98% of students skipping a problem will skip the next

Progress Networks [McBroom et al., ACE'21 10.1145/3441636.3442366]

- Most students fail at least one validation test (most commonly first), but pass after using feedback.

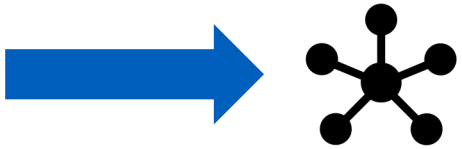
Approx. 10% of first validation test failures are due to syntax errors.

Analysis Approach

Source: Terminal run data


Program code


Error message


Group by error message signature


Manually identify source of error

For Questions 1 & 2

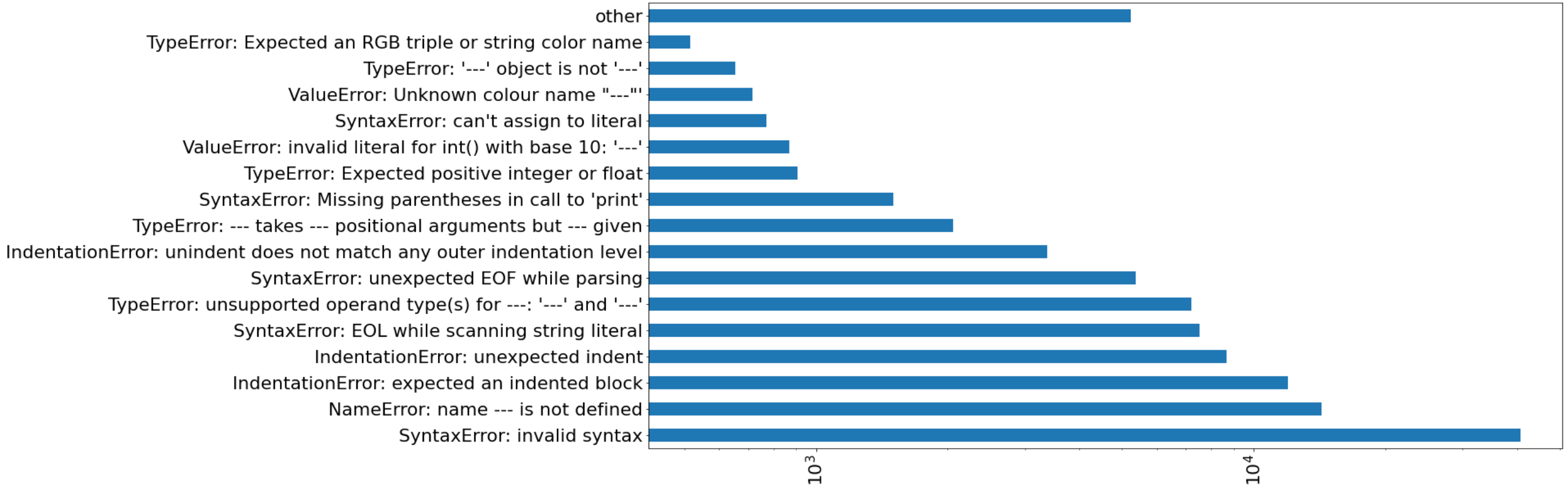

Craft regexes to categorise error


Apply to other terminal runs and aggregate

Q1 & Q2
Deep Dive

Description	Example	Resulting Errors	Students
Newline in string	`print('Hello, World!␣␣␣␣␣')`	`SyntaxError: EOL while scanning string literal`	48
Unmatched quotes	`print('Hello, 'World!')`	`SyntaxError: invalid syntax`	62
	`print('Hello, World!'')`	`SyntaxError: EOL while scanning string literal`	
Mismatched brackets	`print(('Hello, World!')`	`SyntaxError: invalid syntax`	55
Misspelled function name	`PrinT ("One␣fish,␣two␣fish")`	`NameError: name '---' is not defined`	200
Incorrect function name	`write('One␣fish,␣Two␣fish')`	`NameError: name '---' is not defined`	29
Missing parentheses	`print 'One␣fish,␣two␣fish'`	`SyntaxError: Missing parentheses in call to 'print'`	677
	`print'One␣fish,␣two␣fish'`	`SyntaxError: invalid syntax`	559
Accidental whitespace	` print ('One␣fish,␣two␣fish')`	`IndentationError: unexpected indent`	246
Wrong brackets	`print['one␣fish,␣two␣fish']`	`TypeError: 'builtin_function_or_method' object is not subscriptable`	25
Quoted command	`'print' ('One␣fish,␣two␣fish')`	`TypeError: 'str' object is not callable`	7
Incomplete string	`print('One␣fish,␣two␣fish)`	`EOL while scanning string literal`	46
Unquoted string	`print(One fish, two fish)`	`SyntaxError: invalid syntax`	310
	`One fish, two fish`	`SyntaxError: invalid syntax`	214

Whole Course Error incidence

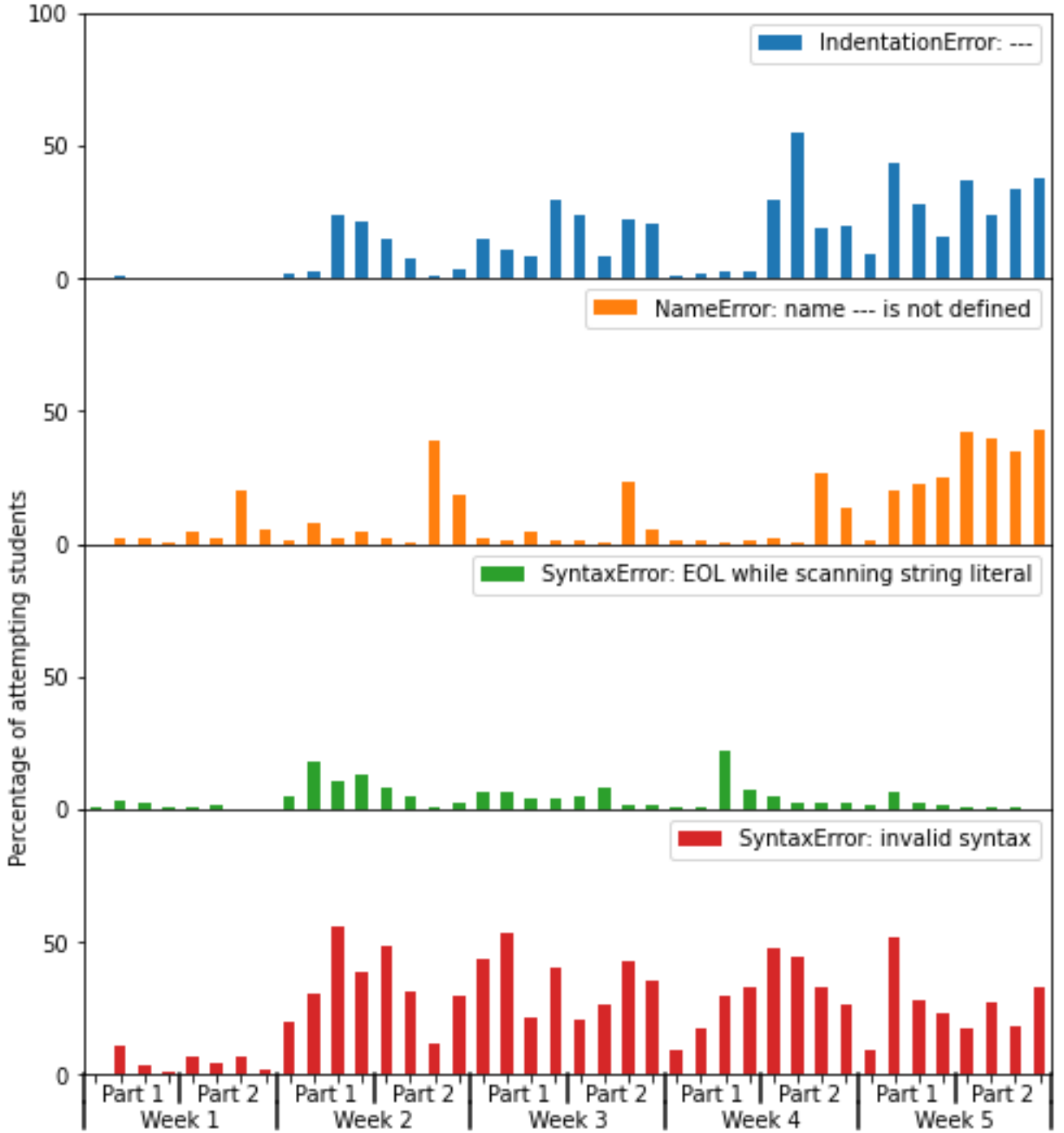


grok academy

Incidence throughout course

Significant programming constructs:

- branching (Week 2)
- loops (Week 3)
- nested conditionals (Week 4)
- functions (Week 5)



grok academy

Reflections

- Manual classification essential to find meaningful results in this kind of data
- Syntax error messages != misconceptions
  - not even 1:many correspondence
- Be mindful that young novice programmers pay less attention to:
  - mathematical concepts,
  - spelling,
  - case and white space semantics
- Relative incidence of errors doesn’t diminish over time
- Students trip up even when we provide lots of help

Further Work

- Data Mining of Syntax Errors in a Large-Scale Online Python Course, Jung et al., AIED2022
- Exploring Student Engagement in an Online Programming Course Using Machine Learning Methods, Polito et al., AIED2022
- Why don’t our instruction slides prevent these errors? When do students engage with them? (Honours project in progress)
- Automate classification of mistakes to provide better feedback

NCSS Challenge Data

- Try the course out at https://grok.ac/iticse2022
- De-identified learning data is not a public dataset, but can be made available to research collaborators under usage agreement
  - Multiple years (since 2013, semiannual since 2020)
  - Multiple streams (newbies, beginners blockly, beginners python, intermediate, advanced)
  - Extensive data (marking submissions, saves, terminal runs, access events)

- Get in touch: bryn.jeffries@grokacademy.org